

AI and Education

Examination of the impact of artificial intelligence on education, including effects at different educational levels, mechanisms and limitations of personalized learning, ethical considerations in AI deployment, debates over AI as teaching assistant versus teacher replacement, challenges to academic integrity posed by generative AI, implications of student datafication and learning analytics, and core components of AI literacy for students and educators.

Ghassem Tofghi

Contents

Overview	4
Learning Objectives	4
Motivation	4
Impact of AI on Education at Various Levels	4
Primary education	4
Secondary education	4
Higher education	5
AI and Personalized Learning	5
Mastery-based progression	5
Bayesian Knowledge Tracing (BKT)	5
Deep Knowledge Tracing (DKT)	6
Content and pathway recommendation	6
Limitations observed in practice	6
Narrow skill-band trapping	6
Short-term engagement bias	6
Reduced confidence on novel tasks	6
Autonomy and agency	6
Transparency and explainability	7
Access and justice	7
AI as Teaching Assistant vs AI as Teacher Replacement	7
AI as teaching assistant	7
AI as teacher replacement	7
Comparison of key dimensions	8
Academic Integrity and Generative AI in Student Work	9
Key generative tools currently used by students	9
Detection challenges	9
Major limitations	9
Detection tools	10
Institutional responses observed since 2023	10
Mandatory disclosure and citation policies	10
Shift toward process-oriented assessment	10
Redesign of summative tasks	10
Introduction of parallel “AI-free” examination variants	11
Development of internal guidelines balancing permissible use	11
Datafication of Students and Learning Analytics	11
Example scenario	12
Data types collected	12
Behavioral data	12
Cognitive data	12
Affective data	13
Metadata	13
Platforms and documented examples	13
Risks and criticisms	13



Proxy-based inequality reinforcement	13
Panoptic effect	14
Commercial extraction	14
Opacity	14
False positives in early-warning systems	14
AI Literacy: What Students and Educators Need to Know	14
For secondary and tertiary students	14
Identifying hallucinations and inaccuracies	14
Probabilistic nature of outputs	14
Training data origins and biases	15
Prompt engineering techniques	15
Assessing generated content	15
Environmental costs of large models	15
For educators	15
Tool alignment with learning outcomes	15
Designing valid assessments in AI-accessible environments	15
Understanding proprietary tool data practices	16
Applying institutional policies and regulations	16
Assessing equity implications of tool recommendations	16
Summary	16



Overview

Artificial intelligence modifies teaching, learning, assessment, and institutional practices across educational levels. This lesson examines applications across primary, secondary, and higher education, mechanisms of personalized learning, ethical challenges, changes in teaching roles, academic integrity issues under generative models, consequences of student datafication, and core components of AI literacy. Key concepts are defined, mechanisms are described, evidence is summarized, and concrete examples are provided with references.

Learning Objectives

- Describe AI system implementation at primary, secondary, and higher education levels with specific examples.
- Explain main algorithmic approaches to personalized and adaptive learning.
- Identify principal ethical concerns in AI-driven education and link them to documented cases.
- Distinguish between AI as teaching assistant and AI as near-complete instructional replacement.
- Analyze how generative language models affect academic integrity norms.
- Define datafication and learning analytics and outline their implications.
- Specify core AI literacy components for students and educators.

Motivation

Heterogeneous student preparation, limited instructional time, unequal access to high-quality resources, and increasing demands for measurable outcomes create structural pressures on educational systems. AI offers scalable automation and data-driven adaptation to address some of these pressures while introducing new issues of dependency, surveillance, bias amplification, and shifts in professional roles.

Impact of AI on Education at Various Levels

Artificial intelligence applications differ substantially across educational stages due to distinct developmental needs, curriculum goals, pedagogical priorities, and assessment practices.

Primary education

Focuses on exploratory interaction, foundational skills, social-emotional development, and early computational thinking.

Examples

- [EMYS social robot](#) supports emotion recognition, facial expression imitation, storytelling activities, and basic language interaction in early childhood settings
- PopBots and Bee-Bot programmable floor robots introduce sequencing, loops, directional commands, and simple algorithmic thinking through physical, play-based activities
- AI Duck multilingual conversational agent facilitates early reading comprehension, vocabulary building, and dialogic reading support in classroom and home environments

Secondary education

Emphasizes domain-specific skill building, conceptual understanding, problem-solving fluency, and preparation for advanced academic study.

Examples



- [Carnegie Learning MATHia](#) / Cognitive Tutor delivers step-by-step individualized tutoring in algebra, geometry, and integrated mathematics with embedded knowledge tracing and cognitive model-based feedback
- [Squirrel AI](#) provides fully individualized mathematics instruction covering complete middle- and high-school curricula through continuous real-time adaptation and mastery-based progression
- [Century Tech](#) generates personalized learning pathways across mathematics, science, English, and other core subjects using diagnostic assessment and adaptive content recommendation

Higher education

Centers on complex reasoning, research skills, academic writing, programming proficiency, critical evaluation, and scalable assessment in large-enrollment courses.

Examples

- [Gradescope](#) applies machine learning to cluster similar handwritten or digital responses, semi-automate grading, and provide consistent feedback in large STEM examinations and assignments
- [OATutor](#) open-source intelligent tutoring system offers adaptive scaffolding, step-by-step guidance, and immediate feedback for introductory statistics, computer science, and related quantitative courses
- ChatGPT and Claude integrations support code debugging, pseudocode generation, literature synthesis, summarization of academic papers, draft feedback, and question generation in large-scale undergraduate and graduate courses

AI and Personalized Learning

Personalized learning systems use learner data to dynamically adjust instructional elements including content difficulty, presentation order, pacing, feedback type, hint level, and scaffolding structure. The goal is to optimize learning efficiency and reduce achievement gaps by tailoring instruction more closely to individual needs than is feasible in traditional whole-class teaching.

Mastery-based progression

Learners must demonstrate consistent proficiency (typically 80–100% correct responses on a set of problems) before advancing to the next topic or level. Incorrect answers automatically trigger review of prerequisite material or additional practice on the same concept.

Example: [Khan Academy](#) structures its content sequences so that a unit only unlocks after the learner achieves a high success rate on prerequisite exercises. Remediation modules are automatically assigned when performance drops below the mastery threshold.

Bayesian Knowledge Tracing (BKT)

BKT maintains a probabilistic estimate of a learner's mastery of each skill (latent knowledge state). After every response, the model updates this probability using Bayes' theorem based on the correctness of the answer and the model's estimated guess, slip, learn, and transit parameters. The system then selects the next problem that maximizes expected learning gain.

Example: [ASSISTments](#) employs BKT variants to decide which mathematics problem to present next and when to provide hints. The platform tracks mastery across hundreds of fine-grained skills and adjusts problem difficulty and hint availability accordingly.



Deep Knowledge Tracing (DKT)

DKT replaces the hand-crafted parameters of BKT with a recurrent neural network (typically LSTM or GRU) that directly models the entire sequence of past responses. The network learns temporal patterns and predicts the probability of correct response on future items, enabling more accurate forecasting of knowledge retention and forgetting over time.

Example: [Duolingo](#) uses a deep knowledge tracing approach to model forgetting curves for vocabulary and grammar items. The system predicts when a learner is likely to forget a specific word or rule and schedules spaced repetition reviews at optimal intervals.

Content and pathway recommendation

These systems employ collaborative filtering, content-based filtering, or hybrid methods to recommend learning resources (videos, articles, exercises, courses) based on the learner's past interactions, performance history, stated preferences, and patterns observed in similar learners.

Example: [Coursera](#) recommends courses, lectures, readings, and practice problems using a combination of user history, completion rates, ratings, and similarity to other learners' trajectories. The recommendation engine helps students discover relevant specializations and supplementary materials.

Limitations observed in practice

Narrow skill-band trapping

Overly narrow adaptation can trap learners in limited skill bands and reduce exposure to transfer problems.

Example: In middle school mathematics platforms, conservative mastery thresholds cause students to remain focused on familiar problem types, leading to reduced performance on cross-domain transfer tasks in longitudinal studies.

Short-term engagement bias

Recommendation systems that optimize short-term engagement metrics may avoid presenting challenging material that supports long-term retention (desirable difficulties).

Example: Language learning applications prioritize high-success-rate items, resulting in decreased long-term recall compared to mixed-difficulty schedules in controlled experiments.

Reduced confidence on novel tasks

Highly adaptive paths sometimes correlate with reduced learner confidence when faced with novel or less-scaffolded tasks.

Example: In introductory programming courses, extended use of hint-heavy adaptive systems is associated with lower self-reported self-efficacy on open-ended projects in surveys from multiple institutions.

Autonomy and agency

Extensive dependence on highly adaptive or directive systems can diminish opportunities for self-regulated learning, metacognitive development, and independent problem-solving.



Transparency and explainability

Many AI systems deliver scores, recommendations, classifications, or feedback without providing interpretable rationales, making it difficult for students, teachers, or administrators to understand or contest decisions.

Access and justice

Advanced AI tools frequently operate behind subscription paywalls, require high-bandwidth connections, or depend on modern devices, creating or reinforcing divides between well-resourced and under-resourced institutions and learners.

AI as Teaching Assistant vs AI as Teacher Replacement

Artificial intelligence can be integrated into educational settings in two fundamentally different roles: as a teaching assistant that augments and supports human instructors, or as a near-complete replacement that attempts to deliver most or all components of instruction autonomously. The assistant role leverages AI strengths in scalability, consistency, and routine handling while preserving human expertise in areas where machines remain limited. The replacement role aims for full automation but faces substantial technical, ethical, and pedagogical barriers.

AI as teaching assistant

In this configuration, AI systems perform supportive functions such as answering factual questions, generating practice exercises, providing immediate feedback on well-defined tasks, summarizing content, or handling administrative duties. Human instructors remain central, overseeing curriculum design, facilitating discussions, offering emotional encouragement, adapting to individual student contexts beyond data patterns, and making ethical judgments. This approach allows teachers to focus on higher-order pedagogical activities while extending their reach to larger or more diverse student groups.

Examples

- [Jill Watson](#) virtual teaching assistant, processes and responds to routine, frequently asked questions in large online computer science courses at Georgia Tech using natural language understanding from IBM Watson technology. The system identifies common queries from past semesters, drafts responses, and escalates complex or novel issues to human TAs, thereby reducing workload on repetitive tasks while maintaining 24/7 availability for students
- [DeepTutor](#) engages students in natural language dialogues during physics problem solving, offering context-specific hints, explanations of misconceptions, and step-by-step guidance based on cognitive models of expertise. Instructors use the system as a supplement for homework or practice sessions, reviewing aggregated performance data to inform class discussions
- [Synthesia](#) creates customizable synthetic avatar videos with accurate lip-sync, multilingual voice synthesis, and gesture animation to deliver standardized explanations, worked examples, or remedial content. Teachers incorporate these as flipped classroom materials or differentiated instruction aids, freeing class time for interactive activities

AI as teacher replacement

Bill Gates predicts that within a decade, AI could replace some teaching functions, based on forecasts of rapid growth in AI's technical power and autonomy. Agentic AI is expected to reach near-human or superhuman problem-solving abilities by 2027–2030, capable of independent decision-making, project

management, and advanced tutoring. While AI can increasingly support instruction, particularly in high school where teaching is largely focused on exam preparation, primary and middle school education remains unlikely to be replaced, as human teachers provide essential social, emotional, and developmental guidance that machines cannot replicate.

The impact of AI on education depends on how teaching and assessment are structured. In high schools, AI tutors could take over routine, test-focused instruction, offering personalized practice, feedback, and simulations. However, fully realizing AI’s potential requires reforming assessment methods away from high-stakes, rote testing toward fostering lifelong learning, critical thinking, teamwork, transdisciplinary skills, and multi-literacies. By integrating AI alongside human teachers, education can be enhanced without losing the social and emotional dimensions critical for younger learners.

This paradigm positions AI as the primary instructional agent, responsible for generating or selecting content, sequencing lessons based on performance data, conducting assessments, delivering feedback, tracking progress, and even attempting basic motivational strategies. Human involvement is minimized to initial setup, occasional oversight, or handling exceptional cases. Proponents argue for radical scalability and cost reduction, while critics highlight deficiencies in relational trust, cultural sensitivity, moral reasoning, and adaptation to unpredictable classroom dynamics.

Examples

- [Squirrel AI](#) operates in select Chinese schools to manage complete middle-school and high-school mathematics curricula. The platform performs diagnostic assessments, creates daily personalized lesson plans, delivers interactive content, evaluates responses in real time, adjusts difficulty levels, and advances students upon mastery demonstration, with teachers serving mainly as facilitators for group activities or behavioral issues.
- Experimental fully automated MOOCs on platforms like [Coursera](#) or [edX](#) employ generative AI to dynamically create lecture scripts, generate varied practice questions with solutions, grade open-ended essays using large language models, provide adaptive feedback, and issue credentials based on performance thresholds, eliminating the need for a dedicated course instructor
- Proposed [Alpha School model](#) conceptualizes AI directing the full K-12 instructional day, including subject-specific tutoring, interdisciplinary project guidance, formative evaluations, and progress reporting. Human staff would focus on social development, extracurriculars, and parental communication rather than core academics

Comparison of key dimensions

Dimension	AI as Teaching Assistant	AI as Teacher Replacement
Primary function	Augments human teaching with targeted support	Replaces most human teaching functions with automated processes
Human role retained	Curriculum design, relational support, complex judgment, motivation, classroom facilitation	Oversight, social-emotional support, exceptional cases, policy compliance
Scalability	High for routine tasks in large cohorts	Very high across entire courses or institutions



Dimension	AI as Teaching Assistant	AI as Teacher Replacement
Current technical feasibility	Widely implemented and reliable in narrow domains	Limited to highly structured subjects; still experimental in broad application
Relational and emotional support	Provided by human instructors	Minimal or absent; simulated at basic levels
Handling ambiguity and values	Handled by human instructors	Severely limited or impossible
Risk of over-reliance	Moderate (supplementary use)	High (core instruction delivered by AI)
Typical deployment scale	Large online courses, tutoring supplements, blended classrooms	Pilot programs, specific subjects in certain regions, automated online courses
Evidence of effectiveness	Positive student perceptions and efficiency gains in studies	Mixed results; gains in access but losses in engagement
Ethical considerations	Lower risk of deskilling; preserves teacher agency	Higher risk of job displacement and reduced human interaction

Academic Integrity and Generative AI in Student Work

Generative large language models enable rapid production of coherent text, code, mathematical derivations, and structured responses with minimal prompting. This capability blurs traditional distinctions between authentic student authorship, original effort, and machine-generated content, challenging conventional definitions of plagiarism, intellectual honesty, and academic evaluation.

Key generative tools currently used by students

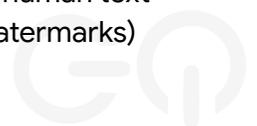
- [ChatGPT](#) — OpenAI’s flagship conversational model, widely used for essay drafting, code generation, and problem solving
- [Claude](#) — Anthropic’s model emphasizing safety and constitutional AI principles, popular for longer-form writing and reasoning tasks
- [Gemini](#) — Google’s multimodal model supporting text, image analysis, and code generation
- [Grok](#) — xAI’s model integrated with real-time web access and designed for helpful, truthful responses
- Locally hosted Llama models (e.g., Llama 3, Llama 3.1) — open-weight models run on personal or institutional hardware, offering privacy and unlimited usage

Detection challenges

Current AI content detectors rely on statistical patterns, perplexity, burstiness, watermarking, or classifier ensembles. Independent evaluations show an overall accuracy of 60–80% under controlled conditions, with significant degradation in real-world scenarios.

Major limitations

- High false-positive rates on non-native English writing and formal academic styles
- High false-negative rates when output is lightly edited, paraphrased, or combined with human text
- Rapid evolution of models that evade detection (e.g., models trained to avoid known watermarks)



- Inability to reliably distinguish between AI-assisted brainstorming and AI-generated final work

Detection tools

Tool	Primary Method	Reported Accuracy (2024–2025 studies)	Key Limitations
Turnitin AI detector	Statistical pattern analysis + proprietary classifier	60–80% (varies by language and editing)	High false positives on non-native English; struggles with short texts or edited AI output
GPTZero	Perplexity and burstiness analysis	70–85% in controlled tests	Less effective on heavily edited or mixed human-AI text
Originality.ai	Ensemble of classifiers + watermark detection	~75–82%	Variable performance on multilingual and technical content
Copyleaks	Deep learning models + linguistic features	~70–80%	Inconsistent results on code and mathematical expressions

Institutional responses observed since 2023

Many universities and school systems have adopted formal policies and assessment redesigns to address the widespread availability of generative language models. Responses fall into several categories: regulation of permissible use, redesign of evaluation methods, and development of new verification practices.

Mandatory disclosure and citation policies

Institutions require students to declare any use of generative AI and to cite the tool and specific prompts employed in the same manner as other sources.

Example

Stanford University policy states that students must explicitly acknowledge AI assistance in written work and include a description of how the tool was used. Failure to disclose constitutes a violation of academic integrity standards

Shift toward process-oriented assessment

Assessment moves from evaluating only the final product to documenting the development process, requiring submission of intermediate artifacts that demonstrate independent thinking.

Examples

- Submission of prompt logs showing iterative interactions with the AI
- Multiple drafts with tracked changes or revision histories
- Annotated bibliographies that separate human-sourced references from AI-generated content
- Reflective statements explaining how AI output was critiqued, modified, or rejected

Redesign of summative tasks

Traditional take-home essays and written reports are replaced or supplemented with formats that are more difficult to complete solely through generative AI.

Examples



- Oral defense or viva voce examination of submitted work
- In-person supervised examinations or timed in-class writing
- Live problem-solving sessions observed by instructors
- Tasks requiring physical demonstration, group negotiation, or real-time data collection

Introduction of parallel “AI-free” examination variants

Some systems maintain two versions of high-stakes assessments: one allowing limited AI use and another explicitly prohibiting all generative tools.

Example

Certain Danish upper-secondary schools have implemented parallel examination formats for select subjects. The AI-free variant uses traditional invigilation and device restrictions to ensure independent work, while the standard version permits defined AI assistance under disclosure rules.

Development of internal guidelines balancing permissible use

Many departments and faculties create subject-specific policies that differentiate between acceptable and unacceptable applications of generative AI.

Examples

- Permitted uses: brainstorming ideas, improving grammar and style, generating reference lists, summarizing readings, checking code syntax
- Prohibited uses: generating complete answers, writing full sections of assessed work, producing original analysis without substantial human contribution
- Conditional uses: allowed only with prior instructor approval, full disclosure, and clear attribution of AI contribution

Broader implications

Generative AI compels institutions to redefine legitimate academic effort away from final-product authenticity toward verifiable demonstration of competence, critical judgment, and intellectual ownership. Assessment increasingly emphasizes:

- Transparency of process
- Ability to critique and refine AI-generated material
- Development of domain-specific reasoning that cannot be fully outsourced
- Cultivation of skills in evaluating the reliability, bias, and limitations of machine-generated content

This shift aligns evaluation more closely with long-term learning goals (independent thinking, ethical scholarship, information literacy) rather than short-term output production.

Datafication of Students and Learning Analytics

Datafication converts diverse student behaviors, responses, interactions, physiological signals, and contextual information into structured, machine-readable datasets. Learning analytics uses statistical, machine learning, and visualization techniques to analyze these data for purposes of description (what happened), diagnosis (why it happened), prediction (what will happen), prescription (what should happen), and intervention (what to do next).



Example scenario

A first-year university student enrolled in an introductory statistics course uses the course learning management system daily. Every click, page view, time spent on video lectures, attempt patterns on quizzes, forum posts, and assignment submission timestamps are automatically recorded. The system also logs response correctness, time taken per question, and hint requests.

The learning analytics platform performs the following:

Description

produces weekly reports showing the student spent 45 minutes on descriptive statistics videos but only 12 minutes on probability exercises.

Diagnosis

identifies that low performance on probability questions correlates with skipping prerequisite algebra review modules.

Prediction

calculates a 72% probability that the student will score below 60% on the upcoming midterm based on current trajectory and historical cohort patterns.

Prescription

recommends that the student complete two specific remedial modules on basic algebra before continuing with probability content.

Intervention

automatically sends the student a personalized email with the recommended links, adds a progress alert to the instructor dashboard, and adjusts the next quiz to include more foundational items.

The instructor reviews the dashboard, sees the student is flagged as moderate risk, and schedules a brief office hours discussion to discuss study strategies.

Data types collected

Behavioral data

Records of observable actions within digital environments.

- Time spent on pages, videos, or exercises (**dwelling time**)
- Click patterns, mouse movements, scroll depth
- Navigation sequences between resources or modules
- Frequency and timing of logins
- **Example platform:** [Century Tech](#) tracks interaction sequences to infer engagement and adapt content delivery

Cognitive data

Measures of knowledge state, skill application, and reasoning processes.

- Response correctness and partial credit patterns
- Hint usage frequency and type
- Error types and systematic misconceptions
- Time to first attempt, time per attempt, revision behavior



- **Example platform:** [ASSISTments](#) logs fine-grained response sequences to update knowledge tracing models and select next problems

Affective data

Indicators of emotional or motivational states (often inferred rather than directly measured).

- Sentiment polarity and intensity in open-text responses or discussion posts
- Webcam-based facial expression analysis (**engagement, confusion, frustration**)
- Keystroke dynamics (typing speed, pauses, corrections) as proxies for confidence or stress
- Self-reported mood scales when explicitly collected
- **Note:** Use of webcam-based affective detection remains controversial due to privacy concerns and limited validity

Metadata

Contextual information attached to primary data streams.

- Timestamp of every logged event
- Device type, operating system, browser version
- Network quality indicators (latency, bandwidth)
- Geolocation (when permitted or inferred)
- Session duration and interruption patterns

Platforms and documented examples

- **Purdue Course Signals** (historical)
Combined LMS activity data (page views, assignment submission patterns, grades) with demographic variables to generate **early-warning risk scores** for dropout or failure. Interventions included targeted emails and advisor meetings.
- **Century Tech**
Continuously collects **behavioral and cognitive data** across subjects to drive **real-time adaptive path adjustment**, content recommendation, and teacher dashboards showing class-level patterns.
- **edX**
Analyzes discussion forum text for **sentiment**, participation frequency, and reply structure. Uses **engagement metrics** (views, posts, upvotes) to predict completion likelihood and inform course design improvements.
- **Other notable systems**
 - **Blackboard Predict:** dropout risk modeling from interaction frequency and assessment patterns
 - **Civitas Learning:** predictive analytics combining behavioral, cognitive, and institutional data
 - **Brightspace (D2L) Student Success System:** real-time risk indicators and recommended actions

Risks and criticisms

Proxy-based inequality reinforcement

Predictive models often rely on proxies that correlate with socioeconomic status rather than controllable academic factors, reinforcing existing inequalities.



Example: Early-warning systems that use login frequency and device type as proxies for engagement frequently assign higher risk scores to students from lower-income households with limited internet access or shared devices, even when academic performance is comparable.

Panoptic effect

Constant monitoring alters student behavior toward compliance rather than genuine exploration.

Example: In platforms that track every click, dwell time, and mouse movement, students report avoiding experimentation with difficult problems or skipping recommended remedial content to maintain high engagement metrics visible to teachers, reducing risk-taking and creative problem-solving.

Commercial extraction

Behavioral surplus collected during learning activities is used for product development, model improvement, or third-party sale.

Example: Several commercial learning platforms have been documented selling anonymized interaction datasets to advertising partners or using them to train general-purpose models unrelated to education, with limited transparency about downstream data flows.

Opacity

Students and teachers rarely understand how scores, risk indicators, or recommendations are computed.

Example: Teacher dashboards displaying “at-risk” flags or suggested interventions often provide no explanation of the underlying algorithm, weightings, or data inputs, leaving educators unable to critically interpret or override system suggestions.

False positives in early-warning systems

Incorrectly labeling students as high-risk can increase stereotype threat or reduce self-efficacy.

Example: In multiple university implementations, students flagged as at-risk based on early LMS patterns (despite later strong performance) experienced lower motivation and higher anxiety, with some studies reporting measurable declines in subsequent assignment completion rates after receiving automated warning messages.

AI Literacy: What Students and Educators Need to Know

AI literacy consists of knowledge and skills that enable informed, critical, and responsible interaction with artificial intelligence systems, particularly those used in educational contexts.

For secondary and tertiary students

Identifying hallucinations and inaccuracies

Identify factual inaccuracies, hallucinations, and fabricated references in model outputs.

Hallucinations refer to content generated by the model that is presented as factual but is entirely invented, unsupported, or false, often appearing plausible and delivered with apparent confidence.

Example: A language model generates a bibliography that includes nonexistent journal articles or assigns incorrect publication years and DOIs to real publications.

Probabilistic nature of outputs

Recognize the probabilistic nature of generative responses and the absence of true understanding.



Example: Identical prompts can produce different answers across runs because token selection follows learned probability distributions rather than deterministic reasoning or retained knowledge.

Training data origins and biases

Understand that training data derives from vast internet corpora and inherits human biases present online.

Example: Image generation models produce gender-stereotyped depictions of occupations (e.g., software engineers predominantly male, teachers predominantly female) because such patterns dominate the web-scraped training images.

Prompt engineering techniques

Apply prompt engineering techniques to increase output quality, reduce ambiguity, and mitigate unwanted patterns.

Example: Including explicit instructions such as “base your answer only on verified information”, “list all sources used”, “present opposing viewpoints”, or “show step-by-step reasoning” often reduces hallucinations and improves logical structure.

Assessing generated content

Assess AI-generated content for factual accuracy, logical coherence, relevance, and completeness.

Example: When an AI summarizes a scientific paper, students must verify whether the summary accurately represents the original methods, results, and limitations, and whether it omits critical caveats or overstates conclusions.

Environmental costs of large models

Recognize the scale of environmental and computational costs of large models (energy consumption, carbon footprint, water usage).

Example: Training one state-of-the-art large language model can require energy equivalent to the annual electricity use of hundreds of households and consume millions of liters of water for data center cooling.

For educators

Tool alignment with learning outcomes

Evaluate whether a specific AI tool aligns with intended learning outcomes and curriculum standards.

Example: Before adopting an AI-powered writing assistant for a first-year composition course, the instructor compares the tool’s feedback focus (grammar, style, structure) against course objectives (developing original argumentation, critical source use, and voice development) to determine whether it supports or risks undermining key competencies.

Designing valid assessments in AI-accessible environments

Design assessments that remain valid in environments where generative AI is accessible (process documentation, oral components, iterative drafts, in-person tasks).

Example: In a literature analysis course, the instructor replaces a single end-of-term essay with a multi-stage submission requiring annotated reading notes, draft outlines, prompt logs showing AI use, and a final oral defense where students explain their reasoning and respond to follow-up questions.



Understanding proprietary tool data practices

Understand data collection, storage, and training practices of proprietary tools and associated privacy implications.

Example: When considering a cloud-based adaptive mathematics platform, the educator reviews the vendor's privacy policy to confirm whether student problem-solving data is used to improve general models, shared with third parties, or retained indefinitely, and discusses consent requirements with school administration.

Applying institutional policies and regulations

Apply institutional policies, national regulations, and ethical guidelines governing AI use in teaching and assessment.

Example: In accordance with university guidelines that require disclosure of AI assistance in all graded work, the instructor creates a clear syllabus statement specifying permitted uses (brainstorming, editing suggestions) and prohibited uses (generating complete answers), and includes a standard citation template for AI contributions.

Assessing equity implications of tool recommendations

Assess equity implications when recommending or requiring specific AI tools (cost, device requirements, internet access, linguistic accessibility).

Example: Before mandating use of a premium grammar and style AI tool in a large introductory course, the instructor checks subscription costs, minimum device specifications, and language model performance across English varieties and non-native speakers to avoid disadvantaging low-income students or multilingual learners.

Summary

Artificial intelligence transforms education through adaptive applications at primary, secondary, and higher levels, personalized learning via mastery progression, knowledge tracing, and recommendation systems, and automated support or instruction. Ethical issues include bias in training data, privacy risks from extensive behavioral tracking, reduced learner autonomy, limited explainability, and unequal access to advanced tools. AI functions as teaching assistants for routine tasks (Jill Watson, DeepTutor, Synthesia) or attempts full instructional replacement (Squirrel AI, automated MOOCs), with the former currently more practical. Generative models (ChatGPT, Claude, Gemini, Grok, Llama) challenge academic integrity by producing near-human work, prompting mandatory disclosure, process-based assessments, and redesigned tasks. Datafication and learning analytics collect behavioral, cognitive, affective, and meta-data streams for prediction and intervention, raising concerns about surveillance and inequity. AI literacy equips students to detect hallucinations, understand biases, and evaluate outputs, while educators must align tools with outcomes, design valid assessments, and address equity. AI delivers efficiency gains in structured domains but requires careful policy and assessment redesign to manage risks to authenticity, equity, privacy, and educational agency.

